Advances in Complex Systems

Dynamic Partnership Formation for Multi-Rover Coordination

MATT KNUDSON

Carnegie Mellon University[*]
matt.knudson@sv.cmu.edu

KAGAN TUMER

Oregon State University[†]
kagan.tumer@oregonstate.edu

Received (received date)
Revised (revised date)

Coordinating multiagent systems to maximize global information collection both presents scientific challenges and provides application opportunities, such as planetary exploration, and search and rescue. In particular, in many domains where communication is expensive because of limited power or computation, the coordination must be achieved in a passive manner, without agents explicitly informing other agents of their states and/or intended actions. In this work, we extend results on such multiagent coordination algorithms to domains where the agents cannot achieve the required tasks without forming teams. We investigate team formation in three types of domains, one where n agents need to perform a task for the team to receive credit, one where there is an optimal number of agents (n) required for the task, but where the agents receive a decaying reward if they form a team with membership other than n, and finally we investigate heterogeneous teams where individuals vary in construction. Our results show that encouraging agents to coordinate is much more successful than strictly requiring coordination. We also show that using objective functions that are aligned with the global objective and locally computable significantly improve over agents using the global objective directly, and that the improvement significantly increases with complexity.

1. Introduction

Coordinating multiple robots to achieve a system-wide objective in an unknown and dynamic environment is critical to many of today's relevant applications, including the autonomous exploration of planetary surfaces and search and rescue in disaster response. In such cases, the environment may be dangerous, uninhabitable to humans all together, or sufficiently distant from central control that response times require autonomous, coordinated behavior.

In general, most multi-robot tasks can be broadly categorized into [13]: (i) tasks where a single robot can accomplish the task, but where having a multi-robot system

[*] NASA Ames Research Park, CMU, Building 23, Moffett Field, CA 94035
[†] Mechanical Engineering, OSU, 204 Rogers Hall, Corvallis, OR 97330

improves the process (for example, terrain mapping or trash collection); and (ii) tasks where multiple robots are necessary to achieve a task (for example to carry an object). In both cases, coordination requires addressing many challenges (low level navigation, high level decision making, inter-rover coordination) each of which requires some degree of information gathering [40]. However, in the first case, a failure of coordination leads to inefficient use of resources, whereas in the second, it leads to a complete system breakdown.

In this work, we focus on problems of the second type, and investigate the robot objective functions that need to be derived for the overall system to achieve high levels of performance. To that end, we investigate the use of difference objective functions to promote team formation [5, 20]. Such objective functions have previously been applied to multiagent coordination problems of the first type. The key contribution of this work is to extend those results to coordination problems of the second type where unless tight coordination among the agents is established and maintained, the tasks cannot be accomplished. We develop teams within the multirover system using passive means (e.g., no explicit coordination directives) through the coupling of the rovers' objective functions.

The application domain we selected is a distributed information gathering problem. First we explore the case where unless a particular point of interest is observed by n rovers within a small amount of time, the point of interest is not considered as observed. Second we explore the case where there is an optimal number of rovers (n) that need to observe a point of interest within the same time window, but where the system receives some value for observations by teams with other than n members. Finally, we construct a system where the individuals are of differing capabilities, and one of each type is needed to provide optimal behavior.

In Section 2 we discuss the rover exploration problem. In Section 3, we present the problem requiring team formation. In Section 4 we present the problem of encouraging rather than requiring team formation, and in Section 5 we present heterogeneous teams with rovers of two types. Finally in Section 6 we discuss the implication of these results and highlight future research directions.

1.1. *Contributions of this Work*

In multiagent domains where the tasks are complex to the point of requiring multiple individuals to participate for completion, there are a number of control approaches. One popular solution is to install explicit coordination directives where individuals identify a task then communicate with other agents to request cooperation and construct a plan to meet the objectives of that task. For example, agent A may observe that task T is incomplete and requires two agents for completion. Agent A then communicates with agent B and requests assistance, perhaps giving agent B the choice to participate or not.

However in many systems, communication among individuals is expensive or not possible, and therefore addressing tasks where partnership is needed is more difficult.

In these cases, the agents only have the capability of observing the behavior of others within the system, and therefore must either develop a specific policy (i.e., always available for assistance or always identify tasks) or adapt their policy dynamically to facilitate what is needed by others within the system. Systems of this type are additionally challenging as individuals can abandon a task in favor of another.

For example, agent A and agent B may be in the process of traveling to observe a point of interest when agent B identifies a more valuable POI nearby. Agent B cannot tell agent A about the new POI, and may abandon agent A to go investigate. Agent A is therefore faced with a decision; should it continue on in the hope that another agent is available, or follow agent B in the hope that it has taken an appropriate course of action? The contribution of the work in this paper targets these situations for investigation. In particular:

- Through the use of neuro-evolutionary learning algorithms, rovers examine others and their environment and must dynamically form partnerships to make observations of points of interest. The environment changes significantly during the learning process, and therefore the rovers are indirectly penalized for establishing permanent partnerships.
- Explicit coordination directives are absent, as is the ability to communicate, therefore partnership formation must be done through the coupling of rover learning objectives. We determine that encouraging partnership formation over a strict requirement is more robust to changes in system parameters and produces better performance.
- Rovers are given different observation capabilities, developing two types within the system. Utilizing an objective that encourages a rover of each type to make an observation it is shown that dynamically formed heterogeneous partnerships are possible through passive means, and the use of difference objectives produce the best performance in congested systems.

These contributions address the advancement of complex multiagent systems to perform tasks when the individuals composing the system have limited capabilities in interacting with each other and the environment. Specifically through the use of simple adaptive algorithms and difference learning objectives the system can develop complex behaviors in furtherance of the performance of robotic exploration domains.

1.2. *Related Work*

Extending single agent learning approaches to multiagent systems presents difficulties in ensuring that the agents not learn a particular task, but a particular task that is beneficial to the overall system. For a small number of agents, the multiagent aspects can be overlooked. But for true learning multiagent systems, new approaches are needed. They include using Markov Decision Processes for online mechanism design [28], developing new reinforcement learning based algorithms [6, 10, 33, 1], or devising agent-specific objective functions [5, 20, 43, 42].

Non-learning approaches to coordination, based on planning, swarms, auctions, and domain specific algorithms have also been investigated. For example, role allocation and plan instantiation have proven successful for large multi-robot systems [44] and planning has been applied to produce a variety of good non-learning techniques in multiagent systems [12]. Swarm techniques are most applicable to very large sets of agents, in particular the use of particle swarm optimization. Topologically independent algorithms have been developed that apply both locally and globally [48] and PSO has interestingly been utilized in mixed signal analysis [37]. Swarm intelligence has also been applied to teams of robots [41], and several successful applications specific to multi-robot coordination include search and rescue [26, 49, 8], robotic soccer [19], mobile sensor networks [16, 30], mine collection [11], and patrol with adversaries [2].

Approaches in development of teams or coalitions within multiagent systems include utilizing inductive logic programming among individuals for path planning [17], forming coalitions [31, 32], and ad-hoc team development [36, 27]. Negotiated learning architectures have been employed in the formation of coalitions as well [34, 14], while the dynamics of market-based coalitions has also been examined [39]. Many interesting applications of partnerships within multiagent systems have been utilized for the demonstration of research in the area, the majority of which being variations on the box pushing domain [45]. Through the use of heterogeneous teams however, role allocation has been applied through the concept of joint intentions [24] and mobile robots have successfully organized an environment of pucks [18].

Biological inspiration is widely researched in the field of multiagent systems due to the many parallels that can be drawn to swarms and collective intelligence in nature [9]. For example, teams have coordinated to efficiently generate paths and explore environments [46, 35], for foraging tasks [38, 15], manufacturing systems have been studied from a natural perspective [7], and biological models have been used in a multi-agent context to allow for self-defined tasks in single robots [47]. Unexpected applications for biologically inspired swarms have emerged as well, such as data harvesting [22] and for software engineering and analysis at NASA [29].

2. Rover Exploration and Coordination

The multi-rover information gathering problem we investigate in this work consists of a set of rovers that must observe a set of points of interest (POIs) within a given time window [20]. The POIs have different importance to the system, and each observation of a POI yields a value inversely related to the distance the rover is from the POI. In addition, and particular to the work presented in this paper, multiple observations of a POI are either required (Section 3) or highly beneficial (Section 4) to the system objective.

2.1. *Rover Capabilities*

Each rover maps its sensor inputs to an x, y motion relative to its current position to select actions. Each rover utilizes a two layer sigmoid activated artificial neural network to perform this mapping.

Each rover uses an artificial neural network, evolving through an evolutionary algorithm, to map its sensor inputs to an x, y motion relative to its current position [21]. In order to perform the mapping, each rover uses a function approximation (implemented by a two-layer feed forward neural network) to perform this mapping. This approach ensures the non-linear mapping capability necessary to perform the task.

The inputs to this function approximator are four POI sensors (Equation 1) and four rover sensors (Equation 2), where x_q^{POI} and x_q^{ROVER} provide the POI and rover "richness" of each quadrant q, respectively, V_j and L_j are the value and location (within quadrant q) of POI j respectively, L_i is the location of the current rover i and $\theta_{j,q}$ is the separation in radians between the POI and the center of the sensor quadrant. The $\delta(\bullet)$ function is the squared Euclidean distance between the two locations down to a minimum distance to prevent dividing by zero, which is 5 units in our case.

$$x_{i,q}^{POI} = \sum_j \frac{V_j}{\delta(L_j, L_i)} \left(1 - \frac{|\theta_{j,q}|}{(\pi/4)}\right) \tag{1}$$

$$x_{i,q}^{ROVER} = \sum_{k,k\neq i} \frac{1}{\delta(L_k, L_i)} \left(1 - \frac{|\theta_{k,q}|}{(\pi/4)}\right) \tag{2}$$

Two outputs from the function approximator indicate the velocity of the rover (in the two axes parallel and perpendicular to the current rover heading). Because the outputs vary from 0 to 1, they are scaled to be between -10 and 10 units. For homogeneous teams the number of hidden units is 20, while the heterogeneous team networks have 32, found through a standard parameter sweep. Finally, the weights of the neural network are initialized randomly to be between $\pm 1/\sqrt{m}$ where m is the number of incoming links to each node, and are adjusted through an evolutionary algorithm [5, 4] (Figure 1).

The evolutionary search algorithm for ranking and subsequently locating successful networks within a population [25, 23, 20] is applied. The algorithm maintains a population of ten networks, utilizes mutation to modify individuals, and ranks them based on a performance metric specific to the domain. The size of the search space varies with the number of network nodes, therefore for the homogeneous case (8 input, 20 hidden, and 2 output nodes) there are 200 parameters, where for heterogeneous rovers (12 input, 32 hidden) we have 424 parameters. A single episode is executed between mutations, and is defined as 60 seconds of operating time. The search algorithm used is shown in Figure 1 which displays the ranking and mutation steps.

```
Initialize $N$ networks at $T = 0$
For $T < T_{max}$ Loop:
    1. Pick a random network $N_i$ from population
       With probability $\epsilon$: $N_{current} \leftarrow N_i$
       With probability $1 - \epsilon$: $N_{current} \leftarrow N_{best}$
    2. Mutate $N_{current}$ to produce $N'$
    3. Control robot with $N'$ for next episode
    4. Rank $N'$ based on performance
       (objective function)
    5. Replace $N_{worst}$ with $N'$
```

Fig. 1. Evolutionary Algorithm: An ϵ-greedy evolutionary algorithm to determine the weights of the neural networks. For all experiments, *epsilon* is set to 0.1. T indexes episodes, N indexes networks with appropriate subscripts, and N' is the mutated network for use in control of the current episode.

In this domain, mutation (Step 2) involves adding a randomly generated number to every weight within the network. This can be done in a large variety of ways, however it is done here by sampling from a random Cauchy distribution [3] of mean 0 and γ of 0.5 where the samples are limited to the continuous range $[-10.0, 10.0]$. Ranking of the network performance (Step 4) is done using a domain specific objective function, and is discussed in the following section.

2.2. *Rover Objectives*

In these experiments, we used three different objective functions [5, 4, 20] to determine the performance of the rover: the system objective function which rates the performance of the full system; a local objective function that rates the performance of a "selfish" rover; and a difference objective function that aims to capture the impact of a rover in the multi-rover system [5]. More precisely, these three functions are:

- The system objective reflects the performance of the full system. Though rovers optimizing this objective guarantees that the rovers all work toward the same purpose, rovers have a difficult time discerning their impact on this function, particularly as the number of rovers in the system increases.
- The local objective reflects the performance of the rover operating alone in the environment. Each rover is rewarded for the sum of the POIs it alone observed. If the rovers operate independently, optimizing this objective would lead to good system behavior. However, if the rovers interact frequently, then each rover aiming to optimize its own local function may

lead to competitive rather than cooperative behavior.

- The difference objective reflects the impact a rover has on the full system [5, 4]. By removing the value of the system objective where rover i is inactive, the difference objective computes the value added by the observations of rover i alone. Because only POIs to which rover i were closest need this difference computed, this objective is "locally" computable in most instances.

Though conceptually the same, the specifics of these objectives are different for each of the problems described in the following sections. Additionally, the rovers are required to make observations within a small amount of time of each other. Therefore, a "team" or partnership is not considered to have formed unless the rovers in that team visit a POI within 5 seconds of one another, or 8% of the total alloted episode time. This is more strict than needed for simple observational tasks, but allows for the support of such tasks as box pushing or stereo observation of a dynamic target.

3. Requiring Team Formation

In the first problem we examine, the rovers need to form teams to perform a task and contribute to the system objective. In this problem, a POI is considered observed only if n rovers visit that POI from within a certain observation distance and within a small time window. Neither the rover, nor the system receive any value unless multiple observations of a POI occur. This problem formulation ensures that the problem is one that cannot be solved by a single rover and that the team formation is essential to the completion of each task.

To formalize this problem, let us first focus on a problem where the observations of the two rovers closest to a POI are tallied. If more than two rovers visit a POI, only the observations of the closest two that occurred within the time window (5 seconds) are considered and their visit distances are averaged in the computation of the system objective (G), which is given by:

$$G(z) = \sum_i \sum_j \sum_k \frac{V_i \; N^1_{i,j} \; N^2_{i,k}}{\frac{1}{2}(\delta_{i,j} + \delta_{i,k})} \tag{3}$$

where V_i is the value of the ith POI, $\delta_{i,j}$ is the closest distance between jth rover and the ith POI, and $N^1_{i,j}$ and $N^2_{i,k}$ determine whether a rover was within the observation distance δ_o and the closest or second closest rover, respectively, to the ith POI:

$$N^1_{i,j} = \begin{cases} 1 & \text{if } \delta_{i,j} < \delta_o \text{ and } \delta_{i,j} < \delta_{i,l} \; \forall l \neq j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and

$$N^2_{i,k} = \begin{cases} 1 & \text{if } \delta_{i,k} < \delta_o \text{ and } \delta_{i,k} < \delta_{i,l} \; \forall l \neq j, k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Determining the order of observation is done two fold. First, the time t_j of the closest rover within the observation distance δ_o is recorded. If the second closest rover k makes its observation within 5 seconds of j, then an observation has occurred and the above calculation is done. This continues outward for each k until the observation distance is exceeded. If no other rover observed within 5 seconds of j, then the second closest is chosen as j, and the process repeats. Therefore, for both $N^1_{i,j}$ and $N^2_{i,k}$ to equal 1, both rovers are the closest two observations made within 5 seconds of each other.

The single rover objective used by each rover only focuses on the value a rover receives for observing a particular POI, and results in:

$$P_j(z) = \sum_i \frac{V_i}{\delta_{i,j}} \qquad \text{if } \delta_{i,j} < \delta_o \tag{6}$$

where notation is the same as above. This objective promotes selfish behavior only, providing a clear, easy-to-learn signal, but one not aligned with the system objective as a whole. This local objective does not take time into consideration, as the rover is considering only its own observation, and therefore is not concerned with arriving in a timely fashion.

Finally, the difference objective for a rover aims to provide system-wide beneficial behavior, while remaining sensitive to the actions of a rover [5]. This difference objective is given by:

$$D_j(z) = \begin{cases} \sum\limits_i \left(\frac{V_i}{\frac{1}{2}(\delta_{i,j}+\delta_{i,k})} - \frac{V_i}{\frac{1}{2}(\delta_{i,j}+\delta_{i,l})} \right) & \text{if } \delta_{i,j}, \delta_{i,k} < \delta_{i,l} < \delta_o \\ \sum\limits_i \frac{V_i}{\frac{1}{2}(\delta_{i,j}+\delta_{i,k})} & \text{if } \delta_{i,j}, \delta_{i,k} < \delta_o, \;\; \delta_{i,l} > \delta_o \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where l is the third closest rover to POI i (meaning that rovers j and k are the closest two for the first two conditionals). The determination of order in both time and space is done here precisely as described above. All three of these objectives were applied for learning in many different situations, though for brevity, only an environment with 50 POIs and 40 rovers (which was representative of the general performance of the objectives) is presented.

Figure 2 shows a schematic of how these objective functions are computed, given that all three rovers are within the observation radius. Only rovers 1 and 2 ($R1$ and $R2$) are taken into consideration when calculating $G(z)$ because their observation distance ($\delta_{1,1}$ and $\delta_{1,2}$) is closer than $R3$ ($\delta_{1,3}$). For $G(z)$, rover 3's observation is discarded. For the difference objective for rovers 1 or 2, rover 3 is taken into consideration. For example, in calculating Equation 7 for $R2$, the first term considers $R1$ and $R2$, where the second term considers $R1$ and $R3$. That is, $R2$ receives the difference between the observation values of $R1$ and $R2$ and the observation values of R1 and R3.

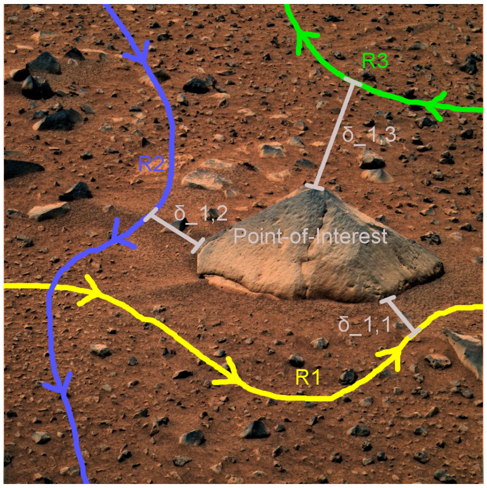

Fig. 2. Sample rover paths in an exploration scenario. Multiple observations are made of a particular point of interest. In the team formation domain, multiple observations must be made for the POI to have any value to the system. Background courtesy of NASA JPL.

3.1. *Results*

The environment used for presentation in this paper contained 40 rovers and 50 POIs, providing a great deal of information to be gathered, while simultaneously creating a congested situation. In addition, the environment was highly dynamic, where 10% of the POIs (selected randomly) changed location and value at each episode. This was done to encourage specific coordination behavior and avoid success of random decisions. The results are based on 3000 episodes of 60 time-steps each, and are averaged over 40 statistical runs for significance, producing the mean and standard error plotted as error bars.

Figure 3 (*left*) plots the performance as the percentage of total value available in the environment. It shows that rovers using all three objectives perform significantly better than random behavior. It also shows that the difference objective provides a signal that allows the rovers to perform better than the local objective. The system objective is too noisy to produce good learning and performs quite poorly. Additionally, Figure 3 (*right*) plots the maximum performance achieved as well as the number of POIs that were fully observed (two valid rover visits) as a percentage of total available. It clearly shows that the difference objective does not contribute to a great deal above system and local until the system reaches the point of high complexity. This is an important conclusion as the work in this paper progresses. Finally, we observe that the percentage of POIs observed is significantly greater than the percentage of POI value, which means that the rovers were actively seeking only to partner, and paid little attention to actual POI value.

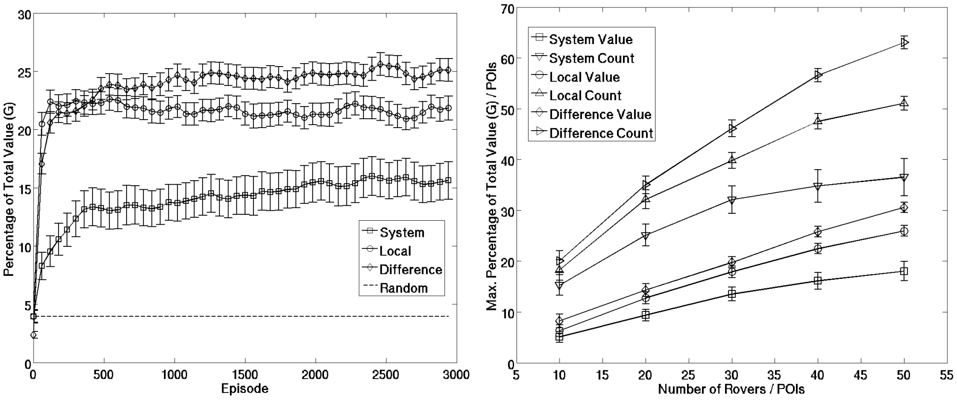

Fig. 3. *Team Formation Required* **Left:** System objective is plotted versus episode as a percentage of the overall value available in the environment for learning in an environment containing 40 rovers and 50 POIs. **Right:** Maximum objective achieved (Value) and number of observed POIs (Count) is plotted for equal numbers of rovers and POIs. Learning is done with system, local, and difference objectives requiring the formation of teams of two rovers.

4. Encouraging Team Formation

In the second problem we examine, multiple rovers are encouraged (rather than required) to form teams to perform a task and contribute to the system objective. In this problem, a POIs value is optimized for n rovers observing it within a specific time window as above, but the system receives lesser value for other numbers of rovers observing the POI. Figure 4 shows the functional form of the two system objectives used in Section 3 and Section 4.

For these objectives, δ_o remains the same, however the distance of observation is no longer explicitly included in the objective, relying on inherent inclusion in the "attendance" to the POI, described as the number of rovers making observations within the alloted time, 5 seconds. This time window could occur at any point during the 60 second episode time, and the determination of order is done using distance first then time as described in Section 3. As before, three objectives are defined, beginning with the system objective given by:

$$G(z) = \sum_i \alpha V_i x e^{\frac{-x}{\beta}} \tag{8}$$

where i indexes POIs, x is the maximum number of rovers within δ_o that made their observations within 5 seconds of one another, β is the observation capacity, and α is a constant chosen to be 2.72 such that the maximum of the exponential curve approximates the POI value V_i.

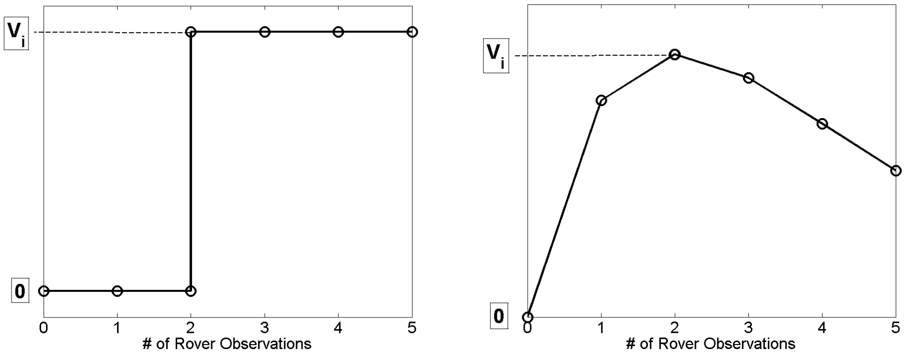

Fig. 4. POI value structure is compared between the required (**left**) and encouraged (**right**) team formation systems.

For this new system objective, the selfish rover objective is defined as:

$$P_j\left(z\right) = \sum_{i_j} \alpha V_{i,j} x e^{\frac{-x}{\beta}} \tag{9}$$

where indexing and constant selection is the same as above. This objective includes no information regarding contribution to the system as a whole, rather indicating only what rover j can directly observe. This rover objective is the component of the system objective for which rover j was within the observation distance δ_o of each POI i.

Finally, the difference objective for this system results in:

$$D_j\left(z\right) = \sum_{i_j} \alpha V_{i,j} \left[x e^{\frac{-x}{\beta}} - \left(x-1\right) e^{\frac{-(x-1)}{\beta}} \right] \tag{10}$$

where indexing and constant selection is the same as above. This objective aims to provide the contribution of rover j to the system. The performance of all three objectives are presented in the next section.

4.1. *Results*

All training parameters were maintained from those used in Section 3.1, including the number of POIs and rovers. The results presented in Figure 5 show a *dramatic* improvement in performance over those shown in 3. First, the percentage of available value in the environment is quite higher (74% versus 25%) for all learning objectives. However, and second, is the much more pronounced performance gain in using the difference objective. The overall performance improvement is explained by the easier learning problem itself, namely by providing a gradient in value obtained, rather than a discontinuous jump. The disparity between the difference objective

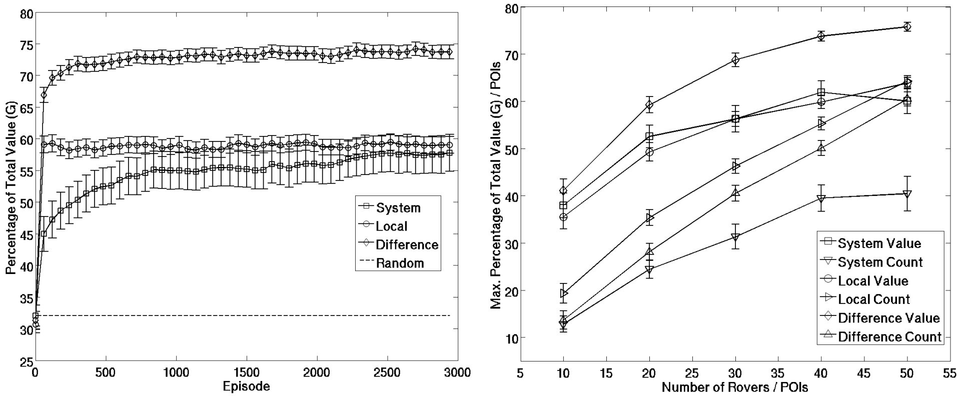

Fig. 5. *Team Formation Encouraged* **Left:** System objective is plotted versus episode as a percentage of the overall value available in the environment for learning in an environment containing 40 rovers and 50 POIs. **Right:** Maximum objective achieved (Value) and number of observed POIs (Count) is plotted for equal numbers of rovers and POIs. Learning is done with system, local, and difference objectives requiring the formation of teams of two rovers.

and the system and local objectives is explained by the clean and accurate learning signal it provides, making it much easier to learn how partnering benefits the system as a whole.

Here again shown in Figure 5 (*right*) that as the system increases in complexity, the difference objective, through providing a better learning signal, provides consistent behavior through the increased complexity of the system. The performance of the system and local objectives falls away sharply compared to the difference, even at moderate complexity, and as expected, the local objective becomes less useful than the system objective at very high complexity because greedy behavior simply will not produce partnerships effectively in large groups. Recalling Figure 3 (*right*) we also observe another sharp contrast, where now the value of the observations far exceeds the number. In fact, the number of observations has remained approximately the same, but the *value* of those observations has dramatically increased.

In encouraging partnership formation, over requiring it, we have presented a simpler problem to learn. This is due to value being assigned for all number of visits to a POI, providing an optimal and therefore a clear gradient to a good solution. As expected then, and shown in Figure 6, the performance of all three objectives has increased. However, the performance gain of the difference objective is pronounced and far exceeds the performance of both the system and local objectives.

4.2. *Higher Coordination Requirements*

The previous two sections investigated coordination for $n = 2$, for both required and encouraged team formation scenarios. The behavior of the system and local objective functions was similar for both cases while the difference objective per-

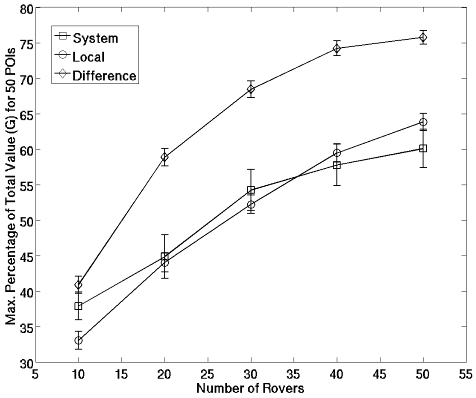

Fig. 6. *Team Formation Encouraged* The maximum system objective achieved is plotted versus varying number of rovers. The number of POIs is held at 50.

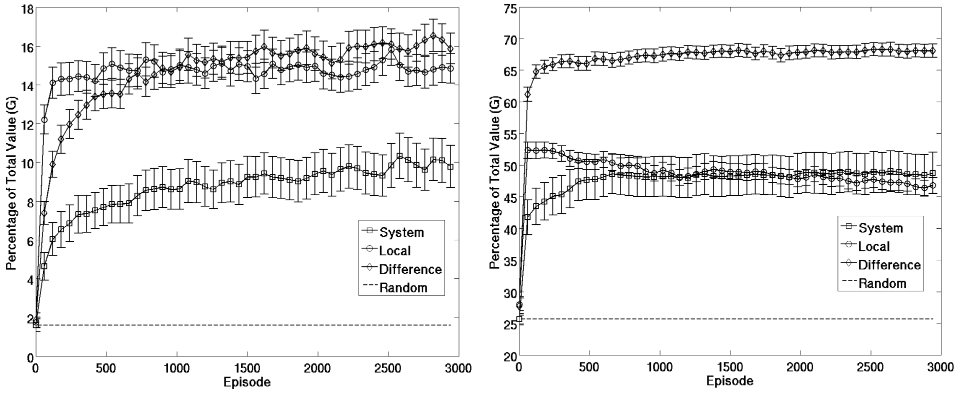

Fig. 7. *Higher Coordination Requirements* ($n = 3$) **Left:** Required Team Formation. **Right:** Encouraged Team Formation. System objective is plotted versus episode for learning in an environment containing 40 rovers and 50 POIs. Learning is done with system, local, and difference objectives for three rovers to observe a POI.

formed quite well in producing good partnerships. In this section we investigate the behavior for $n = 3$, a change that has significant impact on the computation of G, particularly when the observation distance is not increased.

Figure 7 (*left*) shows the learning results for requiring three rovers to observe a POI. The all-or-nothing learning structure in this objective function makes it very difficult for a rover using passive team formation to extract the relevant signal. This brings the difference objective closer to the local objective by reducing its sensitivity to a particular rover's actions (that is, in most cases, removing a rover from the system has no impact on the system performance). As a consequence, the difference objective fails to promote good system-level behavior.

By contrast, Figure 7 (*right*) shows the behavior of the system where team formation is encouraged by a decaying value assignment to POI observations. In this case, moving from $n = 2$ to $n = 3$ does not affect the difference objective. This is because in this problem, removing a rover has a computable impact on the system objective. This creates a "gradient" for evaluating the impact of a rover on the system as a whole. As a consequence, the difference objective performs far better than system or local objectives.

We combine the conclusions that a) encouraging dynamic partnerships, rather than requiring them, is more robust to changes in system definition and, b) difference objectives are more successful in systems changing in the number of rovers and POIs from the above sections to formulate a problem for heterogeneous partnership formation in the following section.

5. Heterogenous Partnership Formation

The success in partnership formation shown in the above sections points to an investigation of teams constructed of heterogenous rovers. When the entire team is made of rovers of identical construction, the tasks are limited to general redundant observations of an environment to provide robustness, or mechanical tasks that require multiple individuals to provide enough effort. In contrast, if the individuals can learn to dynamically partner with one-another, the question arises whether or not, given additional sensing, individuals of differing construction can partner to provide a more specific suite of tasks.

In the final problem we investigate, we define two rover types; *blue* and *green*. These can represent any number of possible construction differences, including sensing and articulation, depending on the system in which they are installed. The individuals must have the ability to determine the difference between the two, for example a blue rover must be able to determine that there are green rovers elsewhere in the environment. In addition, the objective must again be modified to represent the need for rovers of differing capabilities to visit a POI.

The sensing capabilities are similar to those shown in Section 2.1. For each quadrant q however, the rover sensor is split into two, one indicating the density of "blue" rovers and the other indicating "green" rovers. This increases the number of inputs to the neural network from 8 to 12, and the number of hidden units was increased accordingly. This configuration maintains comparability to homogeneous applications while providing the differentiation between rover types needed by the new problem.

We showed that encouraging team formation is more beneficial to the learning process over requiring team formation, and therefore the modified objective reflects the exponential form as much as possible. Again, δ_o remains the same, and the functional form includes the "attendance" of rovers to a given POI. The attendance however is separated into the number of blue rovers and green rovers that made

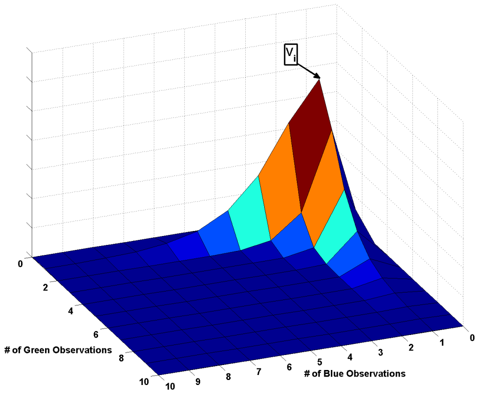

Fig. 8. The objective value is plotted for the number of observations made by blue and green rovers. The maximum POI value V_i is available only when one of each rover type makes an observation.

observations. Similar to Figure 4, Figure 8 shows that the maximum available value V_i for each POI is available only when one rover of each type makes a POI observation. Therefore, the optimal solution is not only that two rovers visit, but that one of each type visits each POI.

As with previous work, three objectives were defined for comparison, reflecting the styles discussed in Section 2.2. Beginning with the system-level objective:

$$G(z) = \sum_{i} \alpha V_i x_{blue} x_{green} e^{\frac{-x_{blue} x_{green}}{\beta_b \beta_g}} \tag{11}$$

where x_{type} is the attendance to POI i of each type of rover, α is a scaling constant to ensure the maximum of the function approximates the POI value V_i (set to 2.72 for these experiments), and β_x are the constants to produce functional peaks at the desired attendance of each type of rover. For example, to have one of each type observe a POI, $\beta_b = \beta_g = 1$, which is the configuration for subsequent experiments.

The local objective is similar to the above, however it reflects only the POIs that rover j has visited. Therefore it is locally computable and easy to learn, but does not indicate the rover's impact on the system as a whole:

$$P_j(z) = \sum_{i_j} \alpha V_{i,j} x_{blue} x_{green} e^{\frac{-x_{blue} x_{green}}{\beta_b \beta_g}} \tag{12}$$

where indexing and constant selection is the same as the above.

Finally, the difference objective includes information contained in the system-level objective, but is easier to learn as it directly indicates how rover j contributed to the system as a whole. It is contingent on the type of rover j:

$$D_j\left(z\right) = \sum_{i_j} \alpha V_{i,j} \left(x_{blue} x_{green} e^{\frac{-x_{blue} x_{green}}{\beta_b \beta_g}} - \left(x_{blue} - 1\right) x_{green} e^{\frac{-\left(x_{blue}-1\right) x_{green}}{\beta_b \beta_g}} \right) \tag{13}$$

where indexing and constant selection is the same as above. The equation shown is for rover j of type *blue*, where if the type is *green*, 1 is subtracted from the *green* rover attendance rather than the *blue*. The experimental results for the use of all three objectives follows in the next section.

5.1. *Results*

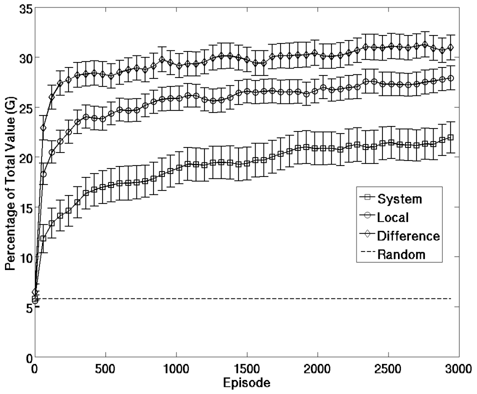

Fig. 9. Heterogeneous Team Formation: System objective as a percentage of total available is plotted versus episode for learning in an environment containing 40 rovers and 50 POIs. Learning is done with system, local, and difference objectives requiring the formation of teams of two rovers, one of each type. The ratio between *blue* and *green* rovers in the system in 50%, providing a balanced team.

The domain for the experiments involving heterogeneous teams is the same as that used in the above work. Each rover is randomly assigned a type at the beginning of each experiment based on a given team ratio. The rovers are still given 60 time-steps for each of 3000 episodes. The environment maintains its dynamic nature, where 10% of the POIs change location and value at every episode, though the rovers maintain their type throughout the learning process.

Shown in Figure 9 are the results of training in an environment where 40 rovers and 50 POIs are present. The ratio of *blue* to *green* rovers is 50%, therefore there are 20 of each type present. With the increased problem complexity we observe that all three learning objectives have significantly decreased performance, even though we are still encouraging partnerships. The difference objective still outperforms however, converging faster and to higher value than the system and local objectives.

As with the results in Section 4.1, learning with the system-level objective proves difficult, as there is a large amount of information contained in the signal; too much regarding other rovers for each individual to ascertain what actions are best in contributing to the system as a whole. The difference objective however, as expected, learns quickly and maintains performance through the learning process. This confirms the applicability of the difference objective in general, and specifically indicates that dynamically requiring heterogeneous team formation in a congested and changing environment is achievable, indeed successful.

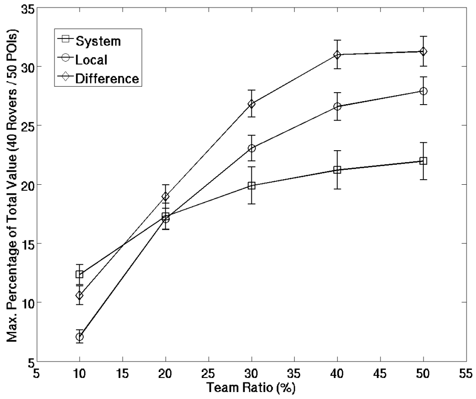

Fig. 10. Heterogeneous Team Ratios: System objective is plotted as a percentage of total available versus episode for learning in an environment containing 40 rovers and 50 POIs. Learning is done with system, local, and difference objectives requiring the formation of teams of two rovers, one of each type. The ratio between *blue* and *green* rovers varies in the system.

In varying the ratio between rover types present in the system, we can determine if the rovers are able to modify their behavior to suit changes in system consistency. For example, if a large set of rovers of a specific type fail, the system must have the ability to adjust coordination behavior to maintain success in accomplishing the tasks requested. Figure 10 shows the maximum achieved system performance as a percentage of total available when the ratio between *blue* and *green* rovers is varied. The variance is symmetrical, therefore 10% *blue* and 90% *green* is the same as 10% *green* and 90% *blue*. The number of rovers and POIs present in the system is held constant.

The system objective in general performs poorly, however the variance in team ratio has a dramatic impact on the difference and local objectives. This is a logical result because as the number of one type of rover goes down, there are far fewer opportunities to partner in the time allotted. The rovers of the minority type must move very quickly, impossibly so the fewer their numbers. As the team becomes balanced then the number of POI observations naturally increases as well. Curiously, the system objective outperforms both the local and difference at 10%, an extremely unbalanced team. One possible explanation for this is that in this extreme the rovers are simply "shooting in the dark", in which case whatever improves the system objective directly (without coherent intention) is the only way to gain any observations at all.

Finally, we sought to determine the impact of varying numbers of rovers and POIs within the system. In many multi-agent problems, the density of agents within the system, as well as the congestion present, can strongly impact the outcome of learning. The above results target a highly dense and congested environment, containing 40 rovers and 50 POIs, where Figure 11 holds each parameter constant and varies the other. The exploration problem is only pertinent when the number of POIs exceeds the number of rovers, therefore the number of rovers is held at a small number (10 in this case) and the number of POIs is varied from 10 to 50. Conversely the number of POIs is held at 50 and the number of rovers within the system varies from 10 to 50. For both investigations the ratio between *blue* and *green* types is 50%.

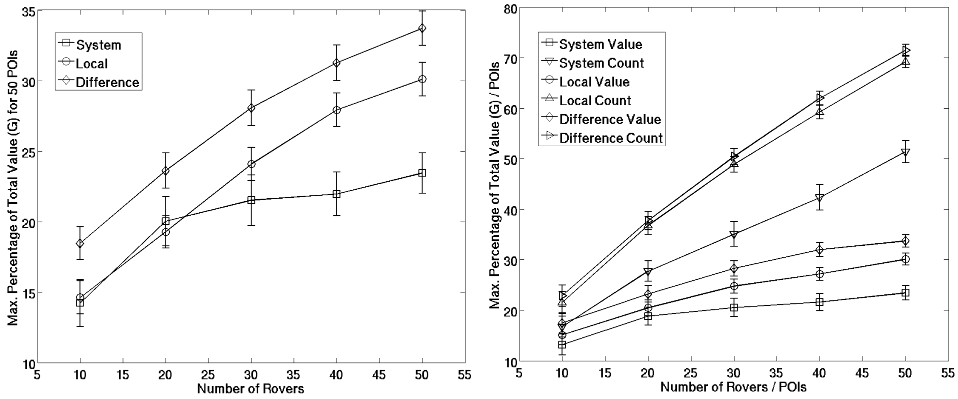

Fig. 11. The maximum system objective achieved as a percentage of the total available is plotted versus varying number of rovers (**left**) and both rovers and POIs (**right**). For the former, the number of POIs is held at 50, and the latter has equal number of rovers and POIs. The team ratio is maintained at 50%.

Figure 11 (*left*) shows the maximum system value achieved as a percentage of the total available. In general it is a classic representation of the differences between the three learning objectives. First we see that in using the system objective, performance can not be maintained as more rovers are present in the system. This again is due to the noise level in the learning signal. For a heterogeneous system, the local objective does surprisingly well and warrants further investigation. One possibility is raised in referencing Figure 11 (*right*) where we see that while the difference objective provides significantly more value than local, the number of POIs visited is very similar. This would suggest that the greedy behavior provided by the local objective still encourages partnership (as nothing would be achieved otherwise), but it does not focus as much on the value of the observations made.

Shown in Figure 11 (*right*) is the maximum system value achieved as a percentage of the total available, as well as the total valid POI observations made (one blue and one green observation). As in Section 3.1 the number of observations again exceeds the value of the observations. Here however this suggests that the rovers are learning first to partner and increasing the value of the observations as a secondary task. Learning parameters here could be modified to adjust for this, whereas in Section 3.1, the objective structure itself produced such aggressive partnering while largely ignoring the value of the observations made.

6. Summary and Future Work

Coordinating multiple robots to achieve a system-wide objective in an unknown and dynamic environment is critical to many of today's relevant applications, including the autonomous exploration of planetary surfaces and search and rescue in disaster response. In this work, we explore multi-robot coordination domains where multiple robots are necessary to achieve a task (for example to carry an object). We focus on passive coordination that is accomplished through the rovers' objective functions.

In all three situations we examined, coordination and team formation is established and maintained through passive means encoded in the rovers objective functions. The difference objective yielded the best results because it provided an objective that was aligned with the overall system objective, while maintaining sensitivity to a rover's actions, even when many rovers were active within the coordinated system. That approach only failed when three or more rovers were required for the completion of a task, without any signal or reward indicating how close to completion that task was. This is an interesting result showing that the difference objective is best suited to domains where the impact of a rover on a system can be ascertained.

We are currently investigating two broad extensions of this work. First, we are investigating the progressive installation of communication capabilities among the individuals within the multi-rover system. Beginning with passive communication (e.g., placement of observation "flags" at a visited POI), capabilities will be increased through additional sensing capabilities (e.g., announced heading to team members), to advanced communication where individuals can send state and intentions to team members. The intention of this extension is to determine how beneficial additional information is for individuals in making decisions not only for exploration but in promoting teams. Second, we are exploring the theoretical basis for the coordination behavior observed in this article. In this extension, we are quantifying the beneficial aspects of coordination as arising through the interactions among the rovers' objective functions. The intent of this work is to lead to rover objective functions that are derived to directly promote coordination without explicit coordination directives.

References

[1] Abdallah, S. and Lesser, V., A multiagent reinforcement learning algorithm with non-linear dynamics, *Journal of Artificial Intelligence Research (JAIR)* **33** (2008) 521–549.

[2] Agmon, N., On events in multi-robot patrol in adversarial environments, in *Int. Conf. on Autonomous Agents and Multiagent Systems* (Toronto, Ontario, 2010), pp. 591–598.

[3] Agogino, A., Tumer, K., and Miikulainen, R., Efficient credit assignment through evaluation function decomposition, in *The Genetic and Evolutionary Computation Conference* (Washington, DC, 2005).

[4] Agogino, A. K. and Tumer, K., Analyzing and visualizing multiagent rewards in dynamic and stochastic environments, *Journal of Autonomous Agents and Multi Agent Systems* **17** (2008) 320–338.

[5] Agogino, A. K. and Tumer, K., Efficient evaluation functions for evolving coordination, *Evolutionary Computation* **16** (2008) 257–288.

[6] Ahmadi, M. and Stone, P., A multi-robot system for continuous area sweeping tasks, in *Proceedings of the IEEE Conference on Robotics and Automation* (Orlando, FL, 2006), pp. 1724–1729.

[7] Barbosa, J., Leitao, P., and Pereira, A., Combining adaptation and optimization in bio-inspired multi-agent manufacturing systems, in *Industrial Electronics (ISIE), 2011 IEEE International Symposium on* (Gdansk, Poland, 2011), pp. 1773 –1778.

[8] Basilico, N. and Amigoni, F., Exploration strategies based on multi-criteria decision making for search and rescue autonomous robots, in *Int. Conf. on Autonomous Agents and Multiagent Systems* (Taipei, Taiwan, 2011), pp. 99–106.

[9] Beckers, R., Holl, O. E., Deneubourg, J. L., universitat Bielefeld, Z., and D-Bielefeld, From local actions to global tasks: Stigmergy and collective robotics, in *Artificial Life IV* (MIT Press, 1994), pp. 181–189.

[10] Busoniu, L., Babuska, R., and De Schutter, B., A comprehensive survey of multiagent reinforcement learning, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* **38** (2008) 156 –172.

[11] Dahl, T., Matari, M., and Sukhatme, G., A machine learning method for improving task allocation in distributed multi-robot transportation, in *Complex Engineered Systems*, eds. Braha, D., Minai, A. A., and Bar-Yam, Y., Vol. 14 (Springer Berlin / Heidelberg, 2006), pp. 307–337.

[12] de Weerdt, M. and Clement, B., Introduction to planning in multiagent systems, *Multiagent and Grid Systems* **5** (2009) 345–355.

[13] Gerkey, B. P. and Mataric, M. J., A formal analysis and taxonomy of task allocation in multi-robot systems, *Robotics Research* **23** (2004) 939–954.

[14] Glinton, R., Scerri, P., and Sycara, K., Exploiting scale invariant dynamics for efficient information propagation in large teams, in *Int. Conf. on Autonomous Agents and Multiagent Systems* (Toronto, Ontario, 2010), pp. 21–30.

[15] Haque, M., Rahmani, A., and Egerstedt, M., Geometric foraging strategies in multi-agent systems based on biological models, in *Decision and Control (CDC), 2010 49th IEEE Conference on* (Atlanta, GA, 2010).

[16] Howard, A., Matari, M. J., and Sukhatme, G. S., An incremental self-deployment algorithm for mobile sensor networks, *Autonomous Robots* **13** (2002) 113–126.

[17] Huang, J. and Pearce, A., Collaborative inductive logic programming for path planning, in *Proc. Int'l Joint Conference on Artificial Intelligence* (Hyderabad, India, 2007), pp. 1327–1334.

[18] Jones, C. and Mataric, M. J., Adaptive division of labor in large-scale multi-robot

systems, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-03)* (Las Vegas, NV, 2003), pp. 1969–1974.

[19] Kalyanakrishnan, S. and Stone, P., Learning complementary multiagent behaviors: A case study, in *RoboCup 2009: Robot Soccer World Cup XIII*, eds. Baltes, J., Lagoudakis, M., Naruse, T., and Ghidary, S., *Lecture Notes in Computer Science*, Vol. 5949 (Springer Berlin / Heidelberg, 2010), pp. 153–165.

[20] Knudson, M. and Tumer, K., Coevolution of heterogeneous multi-robot teams, in *Proceedings of the Genetic and Evolutionary Computation Conference* (Portland, OR, 2010), pp. 127–134.

[21] Knudson, M. and Tumer, K., Adaptive navigation for autonomous robots, *Robotics and Autonomous Systems* **59** (2011) 410–420.

[22] Lee, U., Magistretti, E., Gerla, M., Bellavista, P., Li, P., and Lee, K.-W., Bio-inspired multi-agent data harvesting in a proactive urban monitoring environment, *Ad Hoc Networks* **7** (2009) 725 – 741.

[23] Moriarty, D. and Miikkulainen, R., Forming neural networks through efficient and adaptive coevolution, *Evolutionary Computation* **5** (2002) 373–399.

[24] Nair, R., Tambe, M., and Marsella, S., Team formation for reformation in multiagent domains like robocuprescue, in *RoboCup 2002: Robot Soccer World Cup VI*, Vol. 2752 (2003), pp. 150–161.

[25] Nolfi, S., Floreano, D., Miglino, O., and Mondada, F., How to evolve autonomous robots: Different approaches in evolutionary robotics, in *Proc. of Artificial Life IV* (1994), pp. 190–197.

[26] Nourbakhsh, I., Sycara, K., Koes, M., Yong, M., Lewis, M., and Burion, S., Human-robot teaming for search and rescue, *Pervasive Computing, IEEE* **4** (2005) 72 – 79.

[27] Nouyan, S., Gro, R., Bonani, M., Mondada, F., and Dorigo, M., Teamwork in self-organized robot colonies, in *IEEE Transactions on Evolutionary Computation* (2009), pp. 695–711.

[28] Parkes, D. and Singh, S., An MDP-based approach to online mechanism design, in *NIPS 16* (2004), pp. 791–798.

[29] Pea, J., Rouff, C., Hinchey, M., and Ruiz-Corts, A., Modeling nasa swarm-based systems: using agent-oriented software engineering and formal methods, *Software and Systems Modeling* **10** (2011) 55–62.

[30] Roth, C., Knudson, M., and Tumer, K., Agent fitness functions for evolving coordinated sensor networks, in *Proceedings of the 13th annual conference on Genetic and evolutionary computation* (Dublin, Ireland, 2011), pp. 275–282.

[31] Service, T. and Adams, J., Coalition formation for task allocation: theory and algorithms, *Autonomous Agents and Multi-Agent Systems* **22** (2011) 225–248.

[32] Shrot, T., Aumann, Y., and Kraus, S., On agent types in coalition formation problems, in *Int. Conf. on Autonomous Agents and Multiagent Systems* (Toronto, Ontario, 2010), pp. 757–764.

[33] Singh, A., Krause, A., Guestrin, C., and Kaiser, W., Efficient informative sensing using multiple robots, *Journal of Artificial Intelligence Research (JAIR)* **34** (2009) 707–755.

[34] Soh, L.-K. and Li, X., Investigating adaptive, confidence-based strategic negotiations in complex multiagent environments, *Web Intelligence and Agent Systems* **6** (2008) 313–326.

[35] Sperati, V., Trianni, V., and Nolfi, S., Self-organised path formation in a swarm of robots, *Swarm Intelligence* **5** (2011) 97–119.

[36] Stone, P. and Kraus, S., To teach or not to teach?: decision making under uncertainty in ad hoc teams, in *Int. Conf. on Autonomous Agents and Multiagent Systems*

(Toronto, Ontario, 2010), pp. 117–124.

[37] Sun, T.-Y., Liu, C.-C., Tsai, S.-J., Hsieh, S.-T., and Li, K.-Y., Cluster guide particle swarm optimization (cgpso) for underdetermined blind source separation with advanced conditions, *Evolutionary Computation, IEEE Transactions on* **15** (2011) 798 –811.

[38] Svennebring, J. and Koenig, S., Building terrain-covering ant robots: A feasibility study, *Autonomous Robots* **16** (2004) 313–332.

[39] Tang, F. and Parker, L., A complete methodology for generating multi-robot task solutions using asymtre-d and market-based task allocation, in *Robotics and Automation, 2007 IEEE International Conference on* (Roma, Italy, 2007), pp. 3351 –3358.

[40] Thrun, S. and Sukhatme, G., *Robotics: Science and Systems I* (MIT Press, 2005).

[41] Trianni, V. and Nolfi, S., Engineering the evolution of self-organizing behaviors in swarm robotics: A case study, *Artificial Life* **17** (2011) 183–202.

[42] Tumer, K., Designing agent utilities for coordinated, scalable and robust multiagent systems, in *Challenges in the Coordination of Large Scale Multiagent Systems*, eds. Scerri, P., Mailler, R., and Vincent, R. (Springer, 2005).

[43] Tumer, K., Agogino, A. K., and Welch, Z., Traffic congestion management as a learning agent coordination problem, in *Multiagent Architectures for Traffic and Transportation Engineering*, eds. Bazzan, A. and Kluegl, F. (Springer, 2009), to appear.

[44] Velagapudi, P., Prokopyev, O., Scerri, P., and Sycara, K., A token-based approach to sharing beliefs in a large multiagent team, in *Optimization and Cooperative Control Strategies*, eds. Hirsch, M., Commander, C., Pardalos, P., and Murphey, R., *Lecture Notes in Control and Information Sciences*, Vol. 381 (Springer Berlin / Heidelberg, 2009), pp. 417–429.

[45] Vig, L. and Adams, J., Multi-robot coalition formation, *Robotics, IEEE Transactions on* **22** (2006) 637 –649.

[46] Wagner, I. A., Altshuler, Y., Yanovski, V., and Bruckstein, A. M., Cooperative cleaners: A study in ant robotics, *The International Journal of Robotics Research* **27** (2008) 127–151.

[47] Yu, C.-H. and Nagpal, R., Biologically-inspired control for multi-agent self-adaptive tasks, in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)* (2010), pp. 1702–1707.

[48] Zhan, Z.-H., Zhang, J., Li, Y., and Shi, Y.-H., Orthogonal learning particle swarm optimization, *Evolutionary Computation, IEEE Transactions on* **15** (2011) 832 –847.

[49] Zheng, X., Jain, S., Koenig, S., and Kempe, D., Forest-based multirobot coverage, in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)* (Edmonton, Alberta, 2005), pp. 2318–2323.